Journal of Econometrics 147 (2008) 336–349

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

The structure of US food demand

Jeffrey T. LaFrance [*,1]

School of Economic Sciences, Washington State University, 101 Hulbert Hall, Pullman, WA 99164-6210, United States

ARTICLE INFO

Article history:
Available online 20 September 2008

JEL classification:
D12
E21

Keywords:
Aggregation
Food demand
Functional form
Parameter stability
Rank
Specification errors

ABSTRACT

An exactly aggregable system of Gorman Engel curves for US food consumption is developed and implemented. Box-Cox transformations on prices and income nest functional form. The model nests rank up to rank three. The model is estimated by nonlinear three-stage least squares with annual time series data on 21 foods, 17 nutrients, age and race demographics, and the distribution of income for 1919–1941 and 1947–2000. Results are consistent with full rank three. Point estimates for the Box-Cox parameters on income and prices are 0.86 and 1.09, respectively, strongly rejecting zero and one in both cases. No statistical evidence of serial correlation, specification errors, or parameter instability is found.

1. Introduction

Over the past several years, farm and food policy in the United States has undergone a continuous transformation. The 1996 farm bill replaced farm-level price and income supports with decoupled payments. Welfare, Food Stamps, Women, Infants and Children (WIC), Aid to Families with Dependent Children (AFDC), and School Lunch programs were also reduced in scope and replaced with block grants. Federally subsidized crop insurance has increased from 28 crops on 26 million acres with total liability of $6 billion in 1980 to more than 100 crops on over 220 million acres with total liability in excess of $46 billion in 2006. A current proposal is to expand this program to all forage on public and private grazing land — a net addition of more than 800 million acres. Subsidies for corn ethanol were $7 billion in 2006, leading to significant diversions of wheat, soybean, and other crop acres to corn production in 2007, and rapidly increasing prices for crops and the foods they produce.

These policies and programs and changes in them all influence retail food prices, food quantities consumed, nutrition, food expenditures, and the net incomes of consumers and taxpayers. But we understand poorly their joint effects on the economic well being, food consumption, and nutrition for US consumers. This calls for a coherent, internally consistent model of the demand for food and nutrition. Developing and implementing one such model is the focus of this paper.

As one motivating example among many possible alternatives, consider the joint incentive effects of food stamps and the dairy program. Food stamps provide subsidies for food consumption in an effort to increase the nutritional status of the poor. In contrast, price discrimination in milk marketing orders increase prices paid for fresh milk and lower prices paid for manufactured dairy products (Heien, 1977; Ippolito and Masson, 1978; LaFrance and de Gorter, 1985). These relative price effects cause households to substitute away from fresh milk and towards processed dairy products. Nutritionists and healthcare professionals argue that processed foods containing relatively high levels of fat, cholesterol, salt, sugar, and additives are considerably less healthy than fresh foods that do not contain these factors and are high in fiber, vitamins, and minerals. The upshot is that many farm-level price and income support programs and policies create incentives in direct opposition to those created by food subsidy programs targeting consumers.

The next section of the paper analyzes the theoretical and econometric issues associated with modeling US food consumption and the implied demand for nutrients. Section 3 characterizes the econometric model and its properties. Section 4 discusses the data, empirical results, hypothesis tests, and model diagnostics. Section 5 summarizes and concludes.

2. Modeling food demand

A central focus of a great deal of research on farm and food policy and consumer choice has been an attempt to forge the links

* Tel.: +1 509 335 8512; fax: +1 509 335 1173.
E-mail address: jtlafrance@wsu.edu.

1 I am grateful to the USDA/ERS for their support of this research under cooperative agreement # 43-3AEM-2-80072.

0304-4076/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2008.09.019

between food consumption choices and nutrition. Almost all of the research in this area uses aggregate data, although limited attention has been paid to the implications of aggregation on the structure of economically consistent demand systems. A limited but important and influential subset of the literature on the theory of exactly aggregable demand systems includes: Gorman (1953, 1961, 1965, 1981), Muellbauer (1975, 1976), Deaton (1975, 1986), Howe et al. (1979), Deaton and Muellbauer (1980), Jorgenson et al. (1980, 1982), Russell (1983, 1996), Jorgenson and Slesnick (1984, 1987), Lewbel (1987a,b, 1988, 1989a,b, 1990, 1991, 2003–2004), Diewert and Wales (1987, 1988), Blundell (1988a,b), Wales and Woodland (1983), Brown and Walker (1989), van Daal and Merkies (1989), Jerison (1993), Russell and Farris (1993, 1998), Stoker (1993), Banks et al. (1997), LaFrance et al. (2000, 2002), LaFrance (2004); and LaFrance et al. (2005, 2006).[2]

In this section, I present a method to identify and estimate the impacts of policies on food consumption, nutrition, and consumer welfare using market data by developing and analyzing an exactly aggregable model of US food demand. The purposes of this approach are to derive and implement a theoretically consistent empirical model of household food consumption which: (1) nests the number (i.e., the *rank* of the demand system) and the functional form of the income terms, and the functional form of the price terms in food demand; (2) admits consistent, asymptotically efficient estimation of the food demand parameters with aggregate data; and (3) allows us to draw inferences on the nutritional and welfare effects of farm and food policies on consumers – both in the aggregate and for specific income and other demographic categories.

We first require a fairly large amount of notation. Let $\boldsymbol{p} \in \mathbb{R}_{++}^{n_q}$ be the n_q-vector of market prices for foods, $\boldsymbol{q} \in \mathbb{R}_{+}^{n_q}$, let $\tilde{\boldsymbol{p}} \in \mathbb{R}_{++}^{n_{\tilde{q}}}$ be the $n_{\tilde{q}}$-vector of market prices for other goods, $\tilde{\boldsymbol{q}} \in \mathbb{R}_{+}^{n_{\tilde{q}}}$, let $m \in \mathbb{R}_{++}$ be income, let $s = m - \boldsymbol{p}'\boldsymbol{q} > 0$ be expenditure on other goods, let $\boldsymbol{z} \in \mathbb{R}^K$ be a vector of demographic variables and other demand shifters, let $\pi(\tilde{\boldsymbol{p}})$ be a known, positive-valued, 1° homogeneous, increasing, and weakly concave function of other goods prices, let $\boldsymbol{x} = \left[g_1(p_1/\pi(\tilde{\boldsymbol{p}})) \quad \ldots \quad g_n(p_n/\pi(\tilde{\boldsymbol{p}}))\right]' \equiv \boldsymbol{g}(\boldsymbol{p}/\pi(\tilde{\boldsymbol{p}}))$ be a vector of twice continuously differentiable, strictly increasing functions of deflated prices of foods, and let $y = f(m/\pi(\tilde{\boldsymbol{p}}))$ be a twice continuously differentiable, strictly increasing transformation of deflated income.

Second, we will make extensive use of the real-valued functions,

$$\phi(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x} + 2\boldsymbol{\gamma}'\boldsymbol{x} + 1, \tag{1}$$

$$\theta(\boldsymbol{x}, \tilde{\boldsymbol{p}}, \boldsymbol{z}) = \boldsymbol{\alpha}_0(\tilde{\boldsymbol{p}}, \boldsymbol{z}) + \boldsymbol{\alpha}(\tilde{\boldsymbol{p}}, \boldsymbol{z})'\boldsymbol{x}, \tag{2}$$

where $\boldsymbol{\alpha}(\tilde{\boldsymbol{p}}, \boldsymbol{z})$ is a vector of functions of other prices and demographics, $\boldsymbol{\alpha}_0(\tilde{\boldsymbol{p}}, \boldsymbol{z})$ is a scalar function of other prices and demographics (Pollak and Wales, 1981), both $\boldsymbol{\alpha}(\tilde{\boldsymbol{p}}, \boldsymbol{z})$ and $\boldsymbol{\alpha}_0(\tilde{\boldsymbol{p}}, \boldsymbol{z})$ are 0° homogeneous in $\tilde{\boldsymbol{p}}$, $\boldsymbol{B}$ is an $n_q \times n_q$ matrix of parameters, and $\boldsymbol{\gamma}$ is an n_q-vector of parameters. Due to 0° homogeneity of $\boldsymbol{\alpha}(\tilde{\boldsymbol{p}}, \boldsymbol{z})$ and $\boldsymbol{\alpha}_0(\tilde{\boldsymbol{p}}, \boldsymbol{z})$ in $\tilde{\boldsymbol{p}}$, without any loss in generality, we can (and from this point forward, do) assume that $(\boldsymbol{p}, \tilde{\boldsymbol{p}}, m)$ are deflated by $\pi(\tilde{\boldsymbol{p}})$. From this point forward, I abuse notation slightly and absorb the deflator into the symbols for the price and income variables, so that $(\boldsymbol{p}, \tilde{\boldsymbol{p}}, m)$ denote deflated prices and income.

[2] LaFrance and Pope (2008) synthesize this class of models and extend it to full rank rational demand systems of arbitrary rank in a manner that admits nesting and testing for aggregation, rank, functional form, flexibility, and global regularity of the system of demand equations. Appendix A of the expanded version of this paper discusses the extension of this class of models to incomplete demand systems.

The starting point for the econometric model of US food demand is the class of full rank three exactly aggregable indirect utility functions derived in LaFrance et al. (2005, 2006) and defined by

$$v(\boldsymbol{x}, \tilde{\boldsymbol{p}}, y, \boldsymbol{z}) = \psi\left\{-\frac{\sqrt{\phi(\boldsymbol{x})}}{\left[y - \theta(\boldsymbol{x}, \tilde{\boldsymbol{p}}, \boldsymbol{z})\right]} - \frac{\boldsymbol{\delta}'\boldsymbol{x}}{\sqrt{\phi(\boldsymbol{x})}}, \tilde{\boldsymbol{p}}, \boldsymbol{z}\right\}. \tag{3}$$

Useful choices for the functions $f(\cdot)$ and $\boldsymbol{g}(\cdot)$ are translated Box-Cox transformations, $y = (m^\kappa - 1 + \kappa)/\kappa$ and $x_i = (p_i^\lambda - 1 + \lambda)/\lambda$, $i = 1, \ldots, n_q$. Note that $\kappa = 1$ implies that $y = m$ and $\kappa = 0$ implies that $y = 1 + \ln m$. Analogous relations apply to λ. Thus, this choice produces a demand system that nests the extended price independent generalized linear (PIGL) and the price independent generalized logarithmic (PIGLOG) functional forms (Muellbauer, 1975, 1976). That is, if $\kappa = \lambda = 0$, then we have a full rank three extended translog model (Christensen et al., 1975), while if $\kappa = \lambda = 1$, then we have a full rank three extended quadratic expenditure system (Howe et al., 1979; van Daal and Merkies, 1989). For all values of (κ,λ), we obtain a full rank three quadratic price independent generalized linear (QPIGL) or price independent generalized logarithmic (QPIGLOG) demand system. I call this the *generalized quadratic price independent generalized linear incomplete demand system* (GQ-PIGL-IDS).

A full rank two version of the demand system results when $\boldsymbol{\delta} = \boldsymbol{0}$, while if $\theta \equiv 0$ and $\boldsymbol{\delta} = 0$ then we have a rank one (homothetic) version. Thus, we are able to simultaneously nest rank and functional form of the income terms within a single framework.

Applying Roy's identity gives the demand equations for foods as

$$\boldsymbol{q} = m^{1-\kappa}\boldsymbol{P}^{\lambda-1}\left\{\boldsymbol{\alpha} + \left(\frac{y-\theta}{\phi}\right)(\boldsymbol{B}\boldsymbol{x} + \boldsymbol{\gamma}) + \left[\frac{\boldsymbol{I} - (\boldsymbol{B}\boldsymbol{x} + \boldsymbol{\gamma})\boldsymbol{x}'}{\phi}\right]\boldsymbol{\delta}\frac{(y-\theta)^2}{\phi}\right\}, \tag{4}$$

where $\boldsymbol{P}^{\lambda-1} = \mathbf{diag}\left[p_i^{\lambda-1}\right]$.

This model is nonlinear in income and the demand equations do not aggregate across individuals to per capita income. However, the Gorman class of Engel curves generates theoretically consistent, exactly aggregable models of demand with a small number of statistics concerning the income distribution. In the present case, we need three moments of the income distribution — the cross-section means of $\{m^{1-\kappa}, m, m^{1+\kappa}\}$,

$$\boldsymbol{e} \equiv \boldsymbol{P}\boldsymbol{q} = \boldsymbol{P}^\lambda\left\{\boldsymbol{A}_1(\boldsymbol{x}, \boldsymbol{z})m^{1-\kappa} + \boldsymbol{A}_2(\boldsymbol{x}, \boldsymbol{z})m + \boldsymbol{A}_3(\boldsymbol{x}, \boldsymbol{z})m^{1+\kappa}\right\}, \tag{5}$$

where

$$\boldsymbol{A}_1(\boldsymbol{x}, \boldsymbol{z}) = \boldsymbol{\alpha} - \left(\frac{1+\kappa\theta}{\kappa}\right)(\boldsymbol{B}\boldsymbol{x} + \boldsymbol{\gamma}) + \left(\frac{1+\kappa\theta}{\kappa}\right)^2\left[\frac{\boldsymbol{I} - (\boldsymbol{B}\boldsymbol{x} + \boldsymbol{\gamma})\boldsymbol{x}'}{\phi}\right]\boldsymbol{\delta},$$

$$\boldsymbol{A}_2(\boldsymbol{x}, \boldsymbol{z}) = \left(\frac{\boldsymbol{B}\boldsymbol{x} + \boldsymbol{\gamma}}{\kappa\phi}\right) - 2\left(\frac{1+\kappa\theta}{\kappa^2\phi}\right)[\boldsymbol{I} - (\boldsymbol{B}\boldsymbol{x} + \boldsymbol{\gamma})\boldsymbol{x}']\boldsymbol{\delta}, \quad \text{and}$$

$$\boldsymbol{A}_3(\boldsymbol{x}, \boldsymbol{z}) = \frac{1}{\kappa^2\phi}[\boldsymbol{I} - (\boldsymbol{B}\boldsymbol{x} + \boldsymbol{\gamma})\boldsymbol{x}']\boldsymbol{\delta}.$$

LaFrance et al. (2005, 2006) show that $1 - \boldsymbol{\delta}'\boldsymbol{x}\left(\frac{y-\theta}{\phi}\right) > 0$, $y - \theta < 0$, $\phi > 0$, and $\boldsymbol{B} = \boldsymbol{L}\boldsymbol{L}' + \boldsymbol{\gamma}\boldsymbol{\gamma}'$, where $\boldsymbol{L}$ is lower triangular, are necessary and sufficient for the Slutsky matrix to be symmetric, negative semi-definite in an open neighborhood of $\kappa = \lambda = 1$. The first three conditions are satisfied without imposition for this data set. I impose the system of nonlinear constraints on $\boldsymbol{B} = \boldsymbol{L}\boldsymbol{L}' + \boldsymbol{\gamma}\boldsymbol{\gamma}'$ during estimation (Lau, 1978; Diewert and Wales, 1987).

In the empirical application reported below, the estimated lower triangular matrix $\boldsymbol{L}$ has a reduced rank due to the fact that the